



Deciphering the Drivers of Price-Labor Value Vector Proximity in Empirical Studies

Item Type	Working Paper
Authors	Moraitis, Thanos;Basu, Deepankar
Download date	2026-06-18 15:36:57
Link to Item	https://hdl.handle.net/20.500.14394/59029

Deciphering the Drivers of Price-Labor Value Vector Proximity in Empirical Studies

Thanos Moraitis*

Deepankar Basu†

June 16, 2026

Abstract

In the tradition of Classical Political Economy, empirical studies of labor values and prices ask whether labor-value vectors are close to observed market-price vectors. This paper examines the mechanisms that generate such proximity in input-output systems by uncovering the structural features of the linear production system determining the relevant distance. We use the recent critique of Shaikh-style price-value comparisons by Sabatino (2026) as a springboard for this analysis, by deciphering the results she obtains from different artificial datasets. The central result is that, under the direct-price construction examined here, the distance is governed by the dispersion of reciprocal vertically integrated labor coefficients. This shifts attention to the determinants of the vector of vertically integrated labor coefficients: the dispersion of the direct labor vector, the column structure of the Leontief inverse, and the spectral radius of the input-output matrix. The analysis shows that low market price-labor value distance can arise from identifiable structural mechanisms within the production system.

Keywords: classical theory of prices; labor value; direct prices; input-output analysis; price-value proximity

JEL Codes: B51

1 Introduction

In the Classical Political Economy tradition, labor values measure the direct and indirect labor required to reproduce commodities under given technical conditions. Prices of production are the long-period prices consistent with the equalization of profit rates across

*Department of Economics, University of Massachusetts Amherst. Email: athanasiosmo@umass.edu.

†Department of Economics, University of Massachusetts Amherst. Email: dbasu@umass.edu.

industries, while market prices record the monetary magnitudes at which commodities are actually sold. The relationship between labor values, prices of production, and market prices has therefore occupied a central place in debates over Marxian value theory. One influential strand of the empirical literature, pioneered by Shaikh (1984), has asked whether direct prices, that is, monetary expressions of labor values, are empirically close to market prices and prices of production.

The significance of this empirical claim depends on how the labor theory of value is interpreted. In approaches associated with the New Interpretation, labor values are primarily part of an aggregate accounting framework for analyzing value added, surplus value, and exploitation. On this view, the theory does not require labor values and prices to be close at the level of individual commodities or industries. In much of the empirical price-value literature, however, labor values are treated as gravitational centres for market prices and prices of production. Under this interpretation, the empirical proximity of direct prices, market prices, and prices of production becomes an important part of the case for the continuing relevance of the labor theory of value.

Recent work has reopened this question by emphasizing the sensitivity of price-value deviations to measurement and construction choices. Moraitis and Basu (2026) revisit the empirical literature using both circulating-capital and capital-stock models and find that the evidence for price-value closeness is more mixed than much of the earlier literature suggests. Just as importantly, they show that results depend on how the relevant variables are constructed, including simple labor coefficients, capital stock and depreciation matrices, and price-vector normalizations. This points to a broader methodological lesson: empirical claims about price-value proximity require transparent construction of the underlying price and value vectors.

Sabatino (2026) advances a different critique. She begins by offering the clearest exposition of Shaikh’s empirical procedure, carefully reconstructing the steps through which sectoral gross sales are compared with imputed labor values. On this basis, she argues that the comparison between sectoral gross sales—as measures of prices—and imputed labor values is circular because it compares a monetary magnitude with a function of itself. She further argues that the use of nominal input-output data, rather than physical quantities, creates a gap between the theoretical labor-value system and the empirical test. Finally, she introduces a sequence of artificial datasets, including scrambled sectoral data, random numbers constrained by accounting identities, and a use table modified to spell HUMBUGH with extremely large entries. In each case, she reports a close relationship between gross sales and imputed labor values. She interprets these results as evidence that the method is driven by accounting identities rather than by any substantive relation between labor values

and prices.

This paper uses Sabatino’s critique as an entry point into a broader question: what regulates the observed proximity between market prices and direct prices in empirical input-output systems? Sabatino’s artificial datasets are especially useful for answering this because they make the underlying mechanics clearly visible. They allow us to ask whether proximity is forced by circularity, or whether it arises from identifiable structural features of the constructed linear production system.

Our answer to Sabatino’s three claims is as follows. First, the fact that imputed labor values are constructed from monetary magnitudes does not by itself establish circularity. Functional dependence only shows that two objects are related, not that the distance between them is mechanically forced to be small. Second, Sabatino is right that an empirical procedure based on nominal input-output tables is not identical to a test that could be carried out with physical input-output coefficients. The move to nominal magnitudes changes the space in which the comparison is made, and the resulting distance should therefore be interpreted as a nominal-space comparison rather than as a direct physical-unit comparison. This qualification matters, but it does not make the procedure logically incoherent. Third, Sabatino’s artificial datasets do not show that Shaikh-style calculations are insensitive to economic content in general. They instead reveal specific mechanisms through which constructed input-output systems can lower the market-price/direct-price distance. The relevant task is therefore to identify those mechanisms rather than to treat the results as evidence of generic accounting circularity.

The central result is that, under Sabatino’s direct-price construction, the market-price/direct-price distance is governed by the dispersion of the reciprocal vertically integrated labor coefficients. Let λ denote the vector of vertically integrated labor coefficients. The relevant distance is low when $1/\lambda_i$ is approximately constant across sectors. This shifts the analytical question away from the mere presence of gross output on both sides of the comparison and toward the determinants of dispersion in $1/\lambda$. Our analysis identifies three such determinants: the dispersion of the direct labor vector l , the column structure of the Leontief inverse,¹ and the spectral radius of the input-output matrix.

This perspective changes the interpretation of Sabatino’s artificial datasets. In the Random Numbers case, the fixed income-share construction makes the simple labor vector flat by construction, which compresses the resulting vertically integrated labor coefficients. In the HUMBUG case, the enormous values inserted into the use table push the input-output matrix close to the productivity boundary, so that the Leontief inverse sharply reduces the

¹The Leontief inverse is $(I - A)^{-1}$, where A is the input-coefficient matrix; it records total direct and indirect input requirements.

dispersion of vertically integrated labor coefficients. The low distances in these cases are therefore generated by identifiable features of the constructed systems, not by a general circularity in the comparison between market prices and direct prices.

The paper proceeds as follows. Section 2 situates the argument in the literature on labor values, direct prices, prices of production, and empirical price-value proximity. It also reviews methodological concerns about proximity measures and presents Sabatino's critique as the immediate point of departure. Section 3 re-examines Sabatino's three claims. Section 4 draws out the broader determinants of market-price/direct-price proximity: the dispersion of the direct labor vector, the column structure of the Leontief inverse, and the spectral radius of the input-output matrix. Section 5 concludes.²

2 Background and Literature

2.1 Labor values, direct prices, and the linear production system

The empirical literature on labor values and prices starts from a linear representation of production. Consider an economy with n industries. Let A denote the $n \times n$ input-coefficient matrix, where a_{ij} is the amount of input from industry i required, directly, to produce one unit of gross output in industry j . Let l denote the $1 \times n$ direct labor-coefficient vector, where l_j is the direct labor required per unit of gross output in industry j . In a circulating-capital system, the vector of vertically integrated labor coefficients is then given by

$$\lambda = \lambda A + l \quad \Rightarrow \quad \lambda = l(I - A)^{-1}. \quad (1)$$

where $(I - A)^{-1}$ is the Leontief inverse. The element λ_j measures the total direct and indirect labor required per unit of gross output in industry j .

The corresponding vector of direct prices is obtained by expressing these labor values in monetary terms using a normalization constant. Let Q denote the $n \times 1$ vector of sectoral gross-output monetary magnitudes. The total labor value embodied in the gross output of sector j is

$$\tilde{v}_j = \lambda_j Q_j. \quad (2)$$

The normalization constant is the average labor value of the monetary unit:

$$\alpha = \frac{\sum_{j=1}^n \tilde{v}_j}{\sum_{j=1}^n Q_j}. \quad (3)$$

²The full replication package for the analysis that follows can be downloaded from: <https://drive.google.com/drive/folders/1a-QvORxqPN8P9U6aFPFoLahcHYVpsF4S?usp=sharing>

Direct prices are then constructed by dividing total industry labor values by this average labor value of the monetary unit³:

$$p_j^D = \frac{\tilde{v}_j}{\alpha}. \quad (4)$$

This normalization ensures that aggregate direct prices equal aggregate market-price magnitudes:

$$\sum_{j=1}^n p_j^D = \sum_{j=1}^n Q_j. \quad (5)$$

Market prices, in the empirical setting considered here, are represented by sectoral gross sales or gross output magnitudes, Q_j . Prices of production, by contrast, are long-period prices consistent with the equalization of profit rates across industries.

The empirical literature has therefore typically studied three pairwise relationships: market prices and direct prices, market prices and prices of production, and direct prices and prices of production. Since Sabatino’s critique focuses primarily on the comparison between sectoral gross sales and imputed labor values, the present paper also concentrates on the market-price/direct-price comparison. Prices of production remain central to the broader literature, but they are not the main object of the artificial-data critique examined below⁴.

Once market prices, direct prices, and prices of production have been constructed, the empirical question becomes how their proximity should be measured. Some of the early empirical literature relied on regression-based measures, especially log-log regressions of relative prices on relative labor values, and interpreted high values of R^2 as evidence of price-value proximity. As Basu (2017) highlights, this is problematic because the theory does not imply a high R^2 as such. The relevant test would instead concern whether the regression line has an intercept of zero and a slope of one, so that relative prices and relative values coincide up to random error. For this reason, later work has often turned to direct measures of vector distance.

In what follows, we focus on the unit-independent distance measure discussed by Steedman and Tomkins (1998). Let x denote the column vector of sectoral price-value ratios and let $\mathbf{1}$ denote a vector of ones. Exact equality between prices and values would imply $x = \mathbf{1}$. The angular distance between x and $\mathbf{1}$ is defined as

$$d(x, \mathbf{1}) = \sqrt{2(1 - \cos \theta)}, \quad (6)$$

³Sabatino’s construction differs from the convention used in much of the empirical literature, where direct prices are often compared as unit price vectors rather than total sectoral magnitudes. Since the present paper evaluates Sabatino’s critique on its own terms, we use her construction throughout.

⁴See Moraitis and Basu (2026) for a full treatment of the prices of production system in a circulating and fixed capital model.

where

$$\cos \theta = \frac{x' \mathbf{1}}{\|x\| \|\mathbf{1}\|}. \quad (7)$$

This measure is scale-free and does not depend on the choice of numeraire. The analytical question is therefore what features of the constructed input-output system make the vector x close to $\mathbf{1}$.

A final complication is that these empirical objects are constructed from monetary input-output tables, not directly observed physical input-output coefficients. Empirical labor values therefore depend on how observed nominal data are transformed into the coefficients of the linear production system. We return to this issue in Section 3.

2.2 Empirical evidence on price-value proximity

The modern empirical literature begins with the work by Shaikh (1984) and Ochoa (1984) on the relationship between labor values, prices of production, and market prices, and was developed further by subsequent contributions using input-output data for different countries and periods (for example, see Tsoulfidis and Rieu (2006), Tsoulfidis (2002), and Işıkara and Mokre (2022)). Much of this literature reports high correlations or small distance measures between the relevant price vectors.

In this empirical literature, price-value proximity is generally interpreted as validation of the labor theory of value. Small deviations between direct prices, market prices, and prices of production are taken to show that labor values act as gravitational centres for observed prices. This makes it essential to understand what drives the measured proximity. If low distances can arise from scale effects, low dispersion, normalization choices, or other structural properties of the input-output system, then the theoretical interpretation of those distances becomes less straightforward.

Recent work complicates this picture. Moraitis and Basu (2026) revisits the empirical literature using both circulating-capital and capital-stock models and finds that the evidence for price-value closeness is more mixed and more methodologically fragile than is often acknowledged. In the circulating-capital model, the paper finds substantial cross-country heterogeneity and larger direct-price/market-price deviations than much of the previous literature. In the US capital-stock model, it finds relatively small direct-price/market-price deviations, but much larger deviations between prices of production and both market prices and direct prices. It also shows that these results are sensitive to data construction choices, including the construction of simple labor coefficients, capital stock and depreciation matrices, and price-vector normalizations. The implication is therefore not only that the empirical evidence is mixed, but that claims about price-value closeness require much greater trans-

parency about the construction of the underlying variables.

Earlier critiques had already cautioned against interpreting price-value proximity mechanically. Kliman (2002) argues that high correlations between aggregate sectoral prices and values can reflect industry size rather than a relation between relative prices and relative values. More importantly for the present paper, he shows that small deviations may arise from low dispersion in the data: values can perform little better than a random variable with the same distribution.

This point connects the price-value literature to broader work on regularities in linear production systems. Recent spectral approaches study why empirical input-output systems generate simple patterns, such as near-linear price-profit curves and low-dimensional behavior. Ferrer-Hernández and Torres-González (2021) argue that these regularities depend not only on the eigenvalues of the input-coefficient matrix, but also on the interaction between that matrix and the labor vector through what they call “eigenlabors.”

The present paper is situated in this same methodological space and asks what features of the linear production system mechanically generate market-price/direct-price proximity. This paper contributes to that broader literature by showing that the market-price/direct-price distance is governed by the dispersion of the reciprocal of the vertically integrated labor coefficients.

2.3 Sabatino’s critique and the present paper

Sabatino (2026) advances the most direct recent critique of Shaikh-style empirical tests. The critique has three distinct components. First, Sabatino argues that the comparison between sectoral gross sales and imputed labor values is circular. Since imputed labor values are constructed using monetary magnitudes from the national accounts, the procedure appears to compare gross sales with a function of gross sales, rather than comparing prices with an independently measured quantity of labor. On this interpretation, the test attempts to confirm an accounting identity and is therefore expected to produce results suggesting that labor values and prices are close.

Second, Sabatino argues that Shaikh’s procedure relies on a slippage between the theoretical language of production and the empirical content of the national accounts. The labor-value system is usually presented as if it were built from physical input requirements and direct labor requirements. In practice, however, the input-output tables used in empirical calculation record monetary transactions classified by industry and commodity categories. Sabatino therefore argues that objects described as “inputs,” “outputs,” and “commodities” are not observed physical magnitudes, but accounting categories derived from firms’

monetary records. For Sabatino, this creates a gap between the theoretical object and the empirical test: the theory is formulated in terms of physical production requirements, while the empirical procedure operates on nominal accounting magnitudes.

Third, Sabatino supports the circularity claim through a sequence of artificial datasets. The first scrambles real sectoral data across industries while preserving parts of the accounting structure. The second replaces the underlying system with random numbers constrained by accounting identities. The third modifies a real use table by inserting extremely large entries into selected cells so that the table spells HUMBUG. In each case, Sabatino finds that Shaikh-style calculations continue to produce a close relationship between sectoral gross sales and imputed labor values. She interprets this as evidence that the method is largely insensitive to the economic content of the input data.

These exercises provide the point of departure for the analysis that follows. Section 3 addresses the three claims in turn. It first clarifies why functional dependence does not by itself imply proximity. It then discusses the role of nominal input-output magnitudes in empirical implementations of the labor-value system. Finally, it shows that, under Sabatino's own direct-price construction, the market-price/direct-price distance is governed by the dispersion of vertically integrated labor coefficients. This result allows the artificial datasets to be read not as demonstrations of generic circularity, but as diagnostic experiments that reveal how particular features of constructed input-output systems can generate low distances.

3 Circularity, Nominal Magnitudes, and Artificial Datasets

3.1 Functional dependence and the claim of circularity

Sabatino's first claim is that Shaikh-style tests are circular because sectoral gross sales are compared with imputed labor values that are themselves constructed from gross sales and other monetary magnitudes. This is an important objection, but the conclusion does not follow from the premise. Functional dependence establishes that the two objects are related; it does not establish that the relevant distance between them must be small.

The point is elementary but important. Suppose that a vector x is compared with another vector $f(x)$. This comparison would be circular only if the construction of $f(x)$ forced the relevant distance between x and $f(x)$ to be small. Functional dependence alone does not do this. Even simple transformations can produce vectors that are far from the original vector. Appendix A gives a simple numerical illustration.

The relevant question is therefore more specific. Does the particular function used for direct-price construction mechanically force the market-price/direct-price distance to be

small? The answer depends on the structure of the vector entering the distance statistic. The rest of this section shows that the distance is not automatically small because gross output appears on both sides of the comparison. Under Sabatino's construction, it is governed by the dispersion of vertically integrated labor coefficients. But before we discuss this, let us turn to Sabatino's second point.

3.2 Nominal magnitudes and the empirical labor-value system

Sabatino's second objection concerns the use of nominal input-output data. The theoretical labor-value system is usually written as a physical-unit system. Let A denote the physical input-coefficient matrix, l the vector of direct labor coefficients, and v the vector of labor values. Then

$$v = l(I - A)^{-1}. \quad (8)$$

If the physical coefficients were directly observed, the natural empirical comparison would be between labor values and prices in this physical-unit system.

In practice, however, empirical researchers do not observe the full physical system. Input-output tables record monetary flows between industries. The usual empirical implementation therefore works with a nominally transformed system. Let m denote the vector of market prices and let $M = \text{diag}(m)$. The physical input-output matrix and labor vector are transformed as

$$A' = MAM^{-1}, \quad l' = lM^{-1}. \quad (9)$$

The corresponding transformed labor-value vector is

$$v' = vM^{-1}. \quad (10)$$

This transformation maps the physical labor-value system into the nominal system:

$$v' = l'(I - A')^{-1}. \quad (11)$$

Thus, the use of nominal input-output data is not, by itself, a logical error. It is a way of operationalizing the theoretical system under the data constraints imposed by national accounts.

The important qualification is that this nominal transformation can change the measured distance between two vectors. The distance between two $1 \times n$ vectors, x and y , used in this

paper is based on the angle between the relevant price-value vectors:

$$d(x, y) = \sqrt{2(1 - \cos \theta(x, y))}, \quad \cos \theta(x, y) = \frac{xy'}{\sqrt{xx'}\sqrt{yy'}}. \quad (12)$$

If the move from physical to nominal magnitudes only multiplied every sector by the same scalar, this angle would not change. But the nominal transformation divides each sector by its own market price. In matrix notation, it replaces x and y with xM^{-1} and yM^{-1} .

The distance in the transformed system is therefore

$$d(xM^{-1}, yM^{-1}) = \sqrt{2 \left(1 - \frac{xM^{-2}y'}{\sqrt{xM^{-2}x'}\sqrt{yM^{-2}y'}} \right)}. \quad (13)$$

This expression shows that the nominal-space distance is not generally the same as the physical-space distance. The price matrix changes how sectors enter the angle calculation: sectors with lower market prices receive greater weight, while sectors with higher market prices receive less weight.

Sabatino is therefore right that an empirical test based on nominal input-output data need not reproduce exactly the distance that would be obtained in an unobserved physical-unit system. The point, however, is about metric equivalence, not logical incoherence. The nominal system is a coherent empirical transformation of the physical system, but the resulting distance should be interpreted as a nominal-space distance.

3.3 Artificial datasets and the sources of low distance

Sabatino's empirical challenge rests on a sequence of datasets that are progressively detached from ordinary economic content. The first is a replication dataset based on Shaikh's original exercise. The second scrambles real sectoral magnitudes across industries while preserving parts of the accounting structure. The third constructs a random economy subject to national accounting constraints. The fourth modifies a real use table so that selected cells spell HUMBUG. In each case, Sabatino applies Shaikh's procedure and reports a close relationship between sectoral gross sales and imputed labor values.

Table 1 summarizes the fit statistics reported in Sabatino's figures and adds the angle-based D -distance computed from the corresponding workbooks using Sabatino's direct-price construction. The contrast between the columns is important. Sabatino's reported R^2 and mean absolute error statistics suggest close fit across all four datasets. The scale-free D -distance gives a more differentiated picture: the replication and Scrambled datasets are not especially close relative to many estimates reported in the empirical price-value literature,

while the Random Numbers and HUMBUG datasets generate much lower distances.⁵

Table 1: Fit statistics and angle distances for Sabatino’s datasets

Dataset	Figure	Log-log R^2	Mean absolute error	D -distance
Replication dataset	Figure 4	0.96	14.4%	0.23
Scrambled dataset	Figure 5	0.95	14.8%	0.26
Random Numbers dataset	Figure 6	0.97	8.9%	0.11
HUMBUG dataset	Figure 8	0.99	1.8%	0.09

Notes: The R^2 and mean absolute error columns report the statistics in Sabatino’s figure captions. The D -distance column reports the angle-based distance computed by the authors from the corresponding workbooks using Sabatino’s direct-price construction.

The relevant question is therefore not simply why Sabatino obtains high R^2 values, but why the Random Numbers and HUMBUG constructions reduce the angle-based distance so sharply. Under Sabatino’s direct-price construction, the vector entering the market-price/direct-price distance is proportional to $1/\lambda_i$, where λ_i is the vertically integrated labor coefficient in sector i . If all vertically integrated labor coefficients are equal,

$$\lambda_1 = \lambda_2 = \dots = \lambda_n = \bar{\lambda}, \quad (14)$$

then $1/\lambda_i = 1/\bar{\lambda}$ for every sector. The vector entering the distance calculation is then proportional to a vector of ones. The cosine is equal to one and the angle-based distance is zero.

This limiting case gives the basic intuition. Low distance is produced when $1/\lambda_i$ is nearly constant across sectors; as the dispersion of $1/\lambda_i$ rises, the vector moves away from the constant-vector benchmark and the distance rises. This is why $\hat{C}\hat{V}(1/\lambda)$ is the key diagnostic used below. Appendix B shows formally that the angle-based distance is a monotone function of $\hat{C}\hat{V}(1/\lambda)$. The dataset-specific sections therefore ask how each artificial construction affects the CV of $1/\lambda_i$.

3.3.1 Scrambled: preserving distributional structure while breaking sectoral correspondence

The Scrambled dataset is the least decisive of Sabatino’s artificial examples. Its purpose is to show that the apparent fit between gross sales and imputed labor values survives when sectoral magnitudes are reshuffled across industries. This is a useful diagnostic, but it does not by itself show that the procedure is circular. Scrambling breaks the original sectoral

⁵See Moraitis and Basu (2026) for summary tables of distance measures reported in the empirical literature.

correspondence, but it can still preserve important distributional features of the underlying data. Consistent with this, the angle-based distance *rises* relative to the replication dataset, from 0.23 to 0.26, rather than remaining unchanged.

The more informative exercise is to repeat the scrambling many times and examine what the resulting distances track. We therefore generate repeated scrambled systems and compute, for each valid draw, both the angle-based D -distance and $\hat{C}V(1/\lambda)$. Figure 1 plots the relationship between the two diagnostics. The relationship is effectively one-for-one: draws with higher reciprocal-lambda dispersion have higher D -distances.

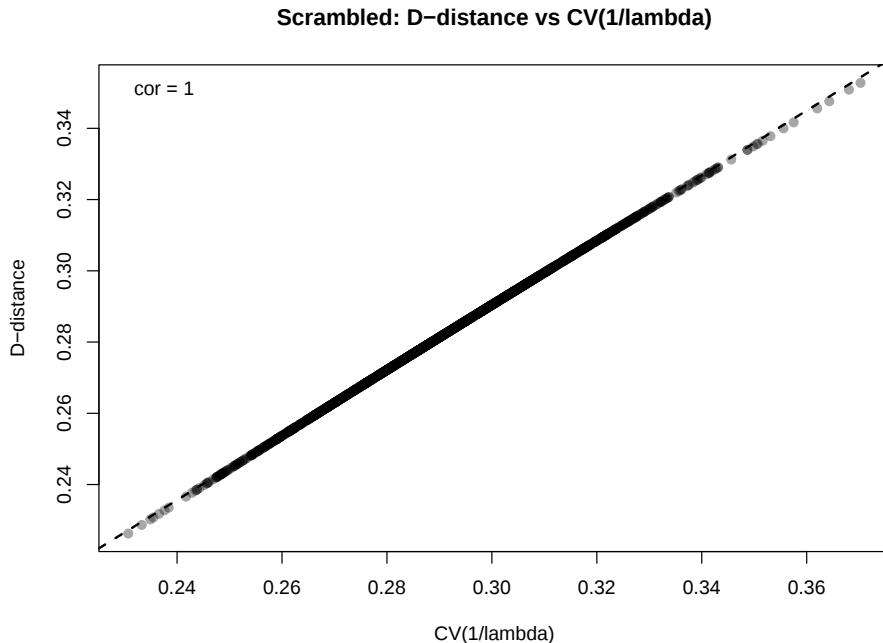


Figure 1: Scrambled draws: D -distance and $CV(1/\lambda)$

The Scrambled dataset therefore mainly illustrates the limits of Sabatino's reported fit statistics. The log-log R^2 remains high, but the scale-free D -distance is not especially low. Repeated scrambling confirms the analytical result above: once Sabatino's direct-price construction is used, the market-price/direct-price distance is governed by the dispersion of the reciprocal vertically integrated labor coefficients.

3.3.2 Random Numbers: fixed income shares and the compression of labor coefficients

The Random Numbers dataset is more informative than the Scrambled dataset because it is not built from reshuffled real sectoral magnitudes. Sabatino constructs an artificial

accounting system from random data. The use table is filled with random numbers; value added is constructed from employment compensation, taxes, and profits; industry gross sales are obtained by adding intermediate output and value added; final uses are normalized to match value added; and commodity gross sales are constructed from the row sums of the use table plus final demand. Sabatino then uses the resulting random accounting system to derive direct-price vectors.

The crucial feature of this construction is how the components of value added are generated. In the Random Numbers dataset, employment compensation, taxes, and profits are fixed proportions of industry intermediate output. Let u_i denote intermediate output in industry i . Sabatino's construction sets

$$c_i = 0.01u_i, \quad t_i = 0.10u_i, \quad \pi_i = 0.15u_i, \quad (15)$$

where c_i is employment compensation, t_i is taxes, and π_i is profits. Value added is therefore

$$y_i = c_i + t_i + \pi_i = 0.26u_i, \quad (16)$$

and industry gross sales are

$$g_i = u_i + y_i = 1.26u_i. \quad (17)$$

This has a direct implication for the simple labor vector. Since the simple labor coefficient is proportional to compensation per unit of gross sales,

$$l_i = \frac{c_i}{g_i} = \frac{0.01u_i}{1.26u_i} = \frac{0.01}{1.26}. \quad (18)$$

The simple labor vector is therefore exactly constant across sectors. This is not a minor detail of the randomization. It is the central mechanism through which the Random Numbers dataset generates a low distance.

To see why, recall that vertically integrated labor coefficients are given in equation 1.

$$\lambda = l(I - A)^{-1}$$

Let $B = (I - A)^{-1}$. If $l_i = \bar{l}$ for all sectors, then

$$\lambda_i = \sum_{j=1}^n l_j B_{ji} = \bar{l} \sum_{j=1}^n B_{ji} = \bar{l} \times \text{sum of } i\text{-th col of } B \quad (19)$$

The vector λ then depends only on the column sums of the Leontief inverse. If those column sums are not highly dispersed, $1/\lambda_i$ will also have low dispersion. Since the market-

price/direct-price distance is a monotone function of $CV(1/\lambda)$, the distance will be low.

We test this mechanism by generating two sets of random accounting systems. The first follows Sabatino’s fixed-share construction. The second preserves the same aggregate income shares but allows sectoral income shares to vary. In the heterogeneous-share version, sector-specific compensation, tax, and profit shares are drawn randomly and then rescaled so that their aggregate shares remain equal to Sabatino’s values, 0.01, 0.10, and 0.15, respectively. The counterfactual therefore preserves the aggregate accounting proportions of the Random Numbers dataset while removing the sector-by-sector equality of income shares. Appendix D.2 describes the construction in detail.

The simulation results are summarized in Figures 2–4. Figure 2 shows the distribution of distances under the fixed-share and heterogeneous-share designs. The fixed-share construction produces distances concentrated near zero, while the heterogeneous-share construction shifts the distribution sharply to the right. Figure 3 shows why: in both designs, the D -distance is governed by $CV(1/\lambda)$, but the heterogeneous-share design generates much higher dispersion in the reciprocal vertically integrated labor coefficients. Finally, Figure 4 reports the same result by design type, making clear that relaxing the fixed-share assumption is sufficient to raise the distance measure substantially.

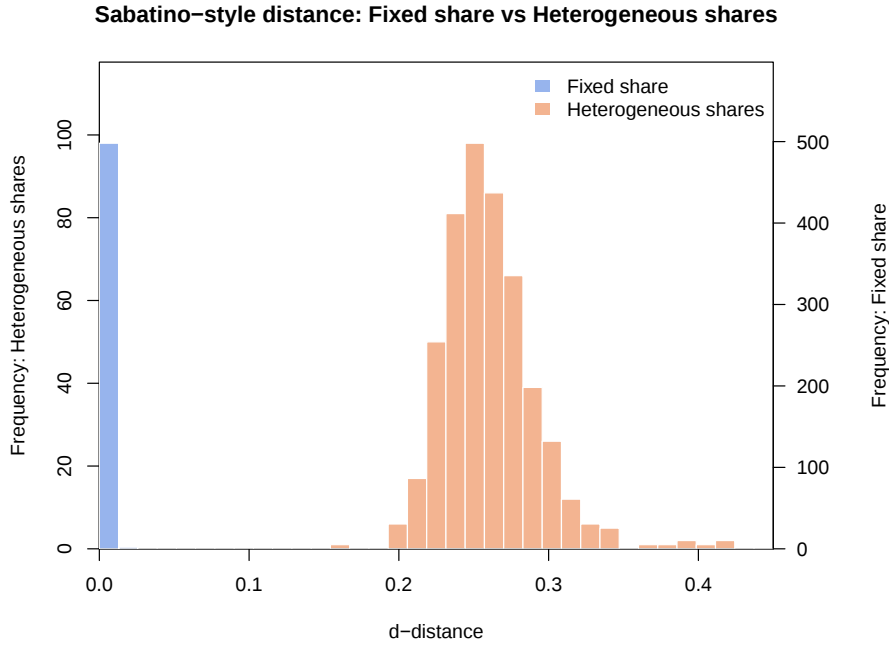


Figure 2: Random Numbers simulations: distance distributions under fixed and heterogeneous shares

The comparison isolates the mechanism. Under fixed sectoral shares, the simple labor

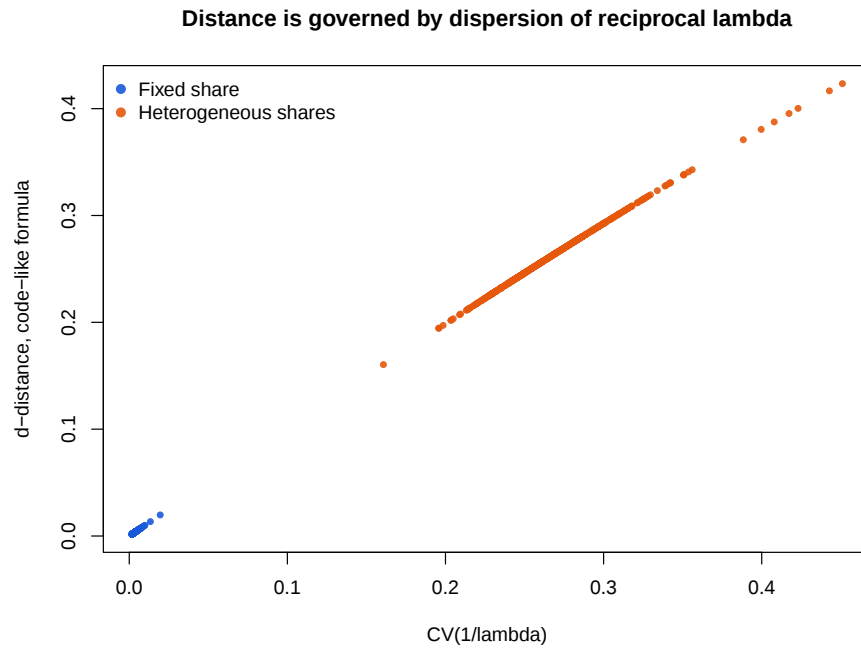


Figure 3: Random Numbers simulations: D -distance and $CV(1/\lambda)$

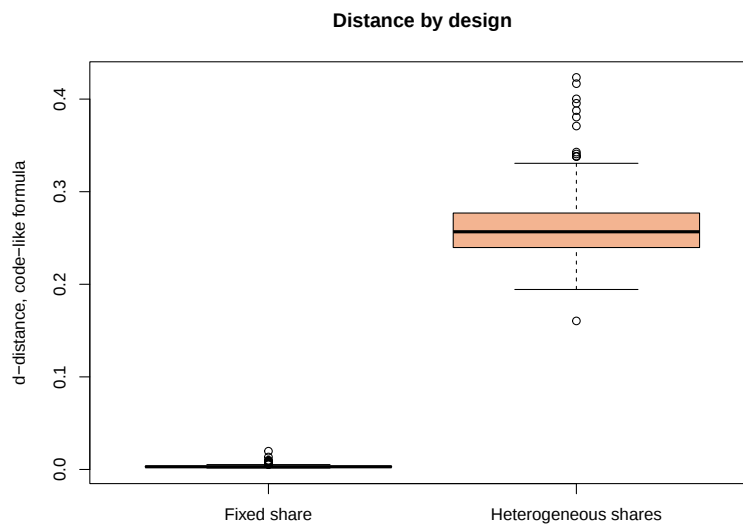


Figure 4: Random Numbers simulations: distance by design

vector is flat by construction, which compresses $CV(1/\lambda)$ and produces a low D -distance. Once sectoral income shares are allowed to vary, while preserving the same aggregate shares, the simple labor vector becomes heterogeneous. This increases the dispersion of vertically integrated labor coefficients and raises the angle-based distance. The Random Numbers result therefore does not show that Shaikh-style calculations produce proximity for arbitrary random data. It shows that a random accounting system with fixed sectoral income shares mechanically compresses the labor vector in a way that lowers the distance statistic.

3.3.3 HUMBUG: spectral radius and the smoothing of vertically integrated labor

The HUMBUG dataset is Sabatino’s most striking artificial example. It begins from an actual use table, but replaces selected cells with the value 11,111,111,111 so that the resulting table visually spells the word HUMBUG. The purpose of the exercise is to show that a visibly artificial and economically implausible input-output table can still generate a close relationship between gross sales and imputed labor values. In Sabatino’s interpretation, this indicates that Shaikh-style calculations are insensitive to the economic content of the input data, much like Shaikh’s own critique of the neoclassical production function (Shaikh 1974).

Our analysis treats the HUMBUG example as a mechanism test. The preceding derivation shows that the market-price/direct-price distance is governed by the dispersion of $1/\lambda$. From 19, the dispersion of λ depends on how the direct labor vector l is filtered through the columns of the Leontief inverse. This suggests two possible channels through which the HUMBUG construction may affect the distance.

The first is a column-structure channel. Since each λ_i is a labor-weighted sum of column i of B , the distance could fall if the HUMBUG entries alter the column structure of the inverse in a way that makes these labor-weighted column sums more similar across sectors. In this interpretation, the distribution of inserted values across columns matters: the word pattern places large entries in particular columns, and this may compress the resulting values of λ_i .

The second is a spectral-radius channel. The HUMBUG entries are not only patterned; they are extremely large. They dominate the scale of the use table and push the input-coefficient matrix A close to the productivity boundary, so that the spectral radius $\rho(A)$ approaches one. As shown formally in Appendix C, when $\rho(A)$ approaches one from below, the Leontief inverse becomes increasingly dominated by the Perron vector of A . In this case, multiplication by B can smooth the vertically integrated labor vector. Since the market-price/direct-price distance is governed by $CV(1/\lambda)$, this smoothing lowers the angle-based distance.

We separate these two channels by constructing two families of HUMBUG variants from

the same pre-HUMBUG use table. In both cases, we keep the same underlying workbook data and the same set of cells that form the HUMBUG letters. The difference is how the artificial entries are assigned to those cells.

The first family tests whether the column pattern of the HUMBUG entries matters. In Sabatino’s workbook, each cell in the HUMBUG mask is assigned the same very large value. The total amount inserted into the use table is therefore this value multiplied by the number of HUMBUG cells. In the column-heterogeneity variants, we keep this total inserted amount unchanged, but redistribute it across the HUMBUG columns. Each column of the HUMBUG mask receives a random multiplier: columns with larger multipliers receive larger inserted values, while columns with smaller multipliers receive smaller inserted values. The inserted values are then rescaled so that the sum of all artificial HUMBUG entries is exactly the same as in Sabatino’s construction. These variants change the distribution of inserted values across columns while holding fixed the overall size of the HUMBUG intervention.

The second family tests whether the overall size of the inserted values matters. Here we keep the HUMBUG mask unchanged and assign the same value to every HUMBUG cell, as in Sabatino’s workbook. Instead of using Sabatino’s original value, however, we choose smaller common values that produce different target values of the spectral radius $\rho(A)$. These variants preserve the visual HUMBUG pattern while changing how close the resulting input-output system is to the productivity boundary. Appendix D.3 describes both constructions in detail.

The results are shown in Figures 5 and 6. Figure 5 compares the workbook HUMBUG result with the generated baseline, the column-heterogeneity variants, and the spectral-radius variants. The column-heterogeneity variants move the distance, but only within a narrow range close to the workbook value. By contrast, the spectral-radius variants span a much wider range of distances. This indicates that the column distribution of the inserted entries matters, but that its effect is small relative to the spectral-radius channel.

Figure 6 isolates the spectral-radius channel directly. Holding the HUMBUG mask fixed, the distance falls as $\rho(A)$ rises. When the inserted values are reduced so that the constructed system is farther from the productivity boundary, the distance is much higher. When the inserted values are large enough to push $\rho(A)$ close to one, the distance falls toward the workbook HUMBUG value. This is the pattern predicted by the spectral-radius mechanism.

The HUMBUG result is therefore not driven simply by the visual fact that the use table spells a word. Nor is it primarily driven by the distribution of inserted values across columns. The dominant mechanism is the scale of the inserted values. Those values push the input-output matrix close to the productivity boundary, change the behavior of the Leontief inverse, smooth the vertically integrated labor coefficients, and lower $CV(1/\lambda)$. Since the

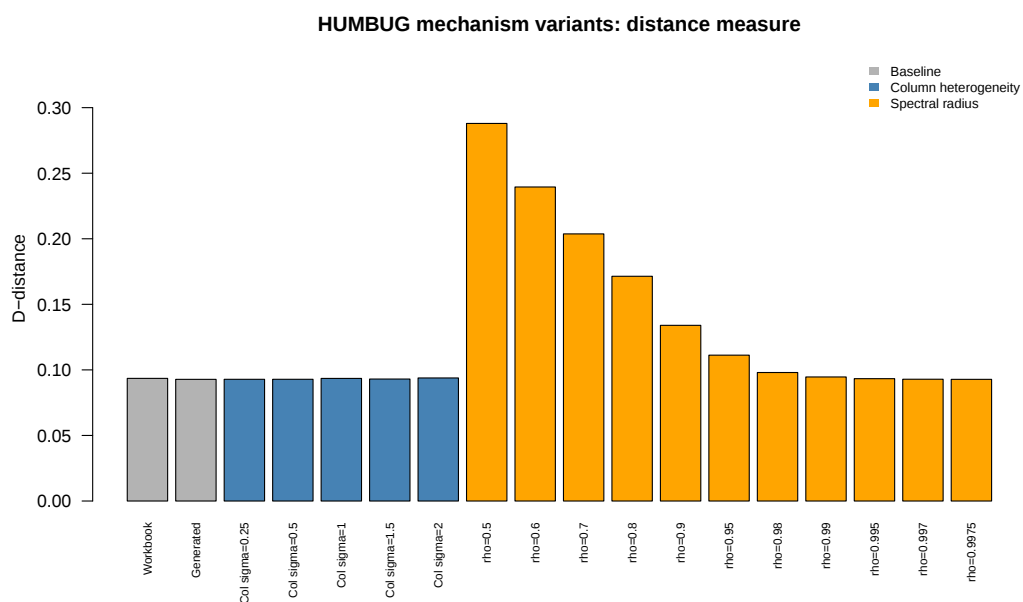


Figure 5: HUMBUG mechanism variants: D -distance

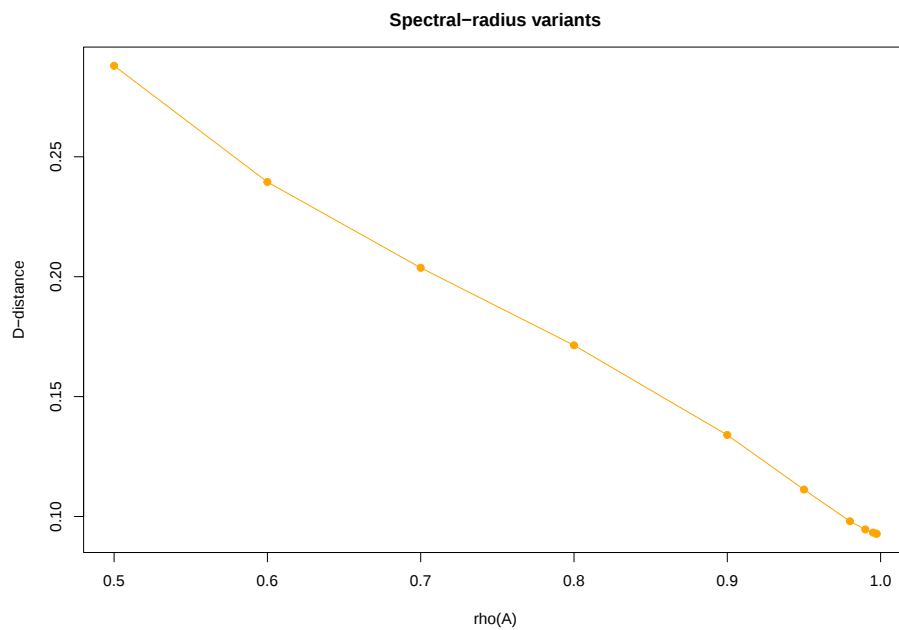


Figure 6: HUMBUG spectral-radius variants: $\rho(A)$ and D -distance

angle-based distance is a monotone function of $CV(1/\lambda)$, the low HUMBUG distance follows from this spectral-radius channel rather than from generic accounting circularity.

4 Discussion

Sabatino’s critique is valuable because it forces a question that the empirical price-value literature has often left implicit: what features of the data and the construction procedure generate the measured proximity between market prices and direct prices? Much of the literature reports correlations, regression fit statistics, or distance measures and interprets them as evidence of price-value closeness. Sabatino shifts attention to the mechanics behind those results. This is the right question to ask, even if the interpretation offered in her paper is not the one supported by the artificial datasets.

A related contribution of Sabatino’s paper is its clarity about how the empirical measures are built from national accounts and input-output data. This clarity is important. As Moraitis and Basu (2026) emphasize, many empirical contributions in this literature do not provide enough detail to assess how the relevant price and value vectors are constructed. Sabatino’s reconstruction of Shaikh’s procedure therefore performs a useful methodological function: it makes the steps of the calculation visible. The present paper builds on that contribution by asking what those steps imply for the distance measure.

The analysis above shows that the circularity claim does not follow from functional dependence alone. To show that a comparison is circular, one would have to show that the construction mechanically forces the relevant distance to be small. Sabatino’s artificial datasets do not show this. Instead, they reveal specific mechanisms through which the constructed system can reduce the dispersion of the vector entering the distance calculation.

The first step is to use the appropriate distance measure. A high log-log R^2 is not a direct test of price-value equality. As Basu (2017) notes, the relevant restriction would be that the regression line has zero intercept and unit slope, not merely that the two variables have high explanatory association. Mean absolute error measures are also sensitive to the scaling used to convert labor values into monetary units. The angle-based D -distance is more appropriate for the present purpose because it is scale-free and directly measures the orientation of the relevant vectors. Even then, the D -distance is not self-interpreting. Its value must be explained.

The key analytical result in this paper is that, under Sabatino’s direct-price construction, the market-price/direct-price distance is a monotone function of $CV(1/\lambda)$. The substantive question is therefore what determines the dispersion of $1/\lambda$ relative to its mean. The analysis identifies three main determinants: the dispersion of the direct labor vector l , the column

structure of the Leontief inverse, and the spectral radius of the input-output matrix.

The first determinant is the dispersion of l . If the direct labor vector is compressed, the resulting vertically integrated labor coefficients are predisposed to be compressed as well. This is the mechanism isolated by the Random Numbers dataset. In Sabatino’s construction, compensation, taxes, and profits are fixed proportions of intermediate input in every sector. This makes the simple labor coefficient constant across sectors by construction. When this fixed-share assumption is relaxed, while preserving the same aggregate income shares, the simple labor vector becomes heterogeneous and the distance rises. The Random Numbers result therefore shows that a random system with fixed sectoral income shares compresses the labor side of the calculation.

The second determinant is the column structure of the Leontief inverse. Since each vertically integrated labor coefficient is a labor-weighted sum of a column of $B = (I - A)^{-1}$, variation in λ_i depends on how the direct labor vector is filtered through the columns of B . Differences in column sums, or more precisely in labor-weighted column sums, can therefore affect the dispersion of λ . This is why column structure is a natural candidate mechanism in the HUMBUG case. The diagnostic variants suggest that this channel is present but limited: redistributing the same total inserted amount across columns increases the D -distance, but the magnitude of this effect is small compared with the changes generated by varying the spectral radius.

The third determinant is the spectral radius of A . This is the dominant mechanism in the HUMBUG example. The values inserted to form the HUMBUG letters are enormous. They push the input-coefficient matrix close to the productivity boundary, raising $\rho(A)$ close to one. In that region, the Leontief inverse smooths the vertically integrated labor vector. When the same HUMBUG mask is retained but the inserted values are scaled down, $\rho(A)$ falls, $CV(1/\lambda)$ rises, and the D -distance increases. The low HUMBUG distance is therefore not primarily a consequence of the word pattern or of the column distribution of the inserted values. It is mainly a consequence of the scale of those values and their effect on the spectral properties of the constructed input-output system.

5 Conclusion

This paper has used Sabatino’s critique as a way to examine what drives market-price/direct-price proximity in empirical input-output systems. The central point is that the appearance of proximity is not self-explanatory. It depends on the construction of the relevant vectors, the metric used to compare them, and the structural properties of the linear production system.

Our analysis yields two main conclusions. First, Sabatino’s circularity claim does not follow from the fact that imputed labor values are constructed using monetary magnitudes. Functional dependence is not sufficient to imply proximity. Nor does the use of nominal input-output data make the procedure logically incoherent, although it does mean that the resulting distance should be interpreted as a nominal-space comparison rather than as a direct physical-unit test.

Second, under Sabatino’s direct-price construction, the market-price/direct-price distance is governed by the dispersion of $1/\lambda$ relative to its mean, where λ is the vector of vertically integrated labor coefficients. This result turns the artificial datasets into diagnostic exercises. The Random Numbers case shows how fixed sectoral income shares can flatten the direct labor vector and reduce the distance. The HUMBUG case shows how extremely large inserted values can push the input-output matrix close to the productivity boundary, smoothing vertically integrated labor coefficients through the Leontief inverse. In both cases, low distance arises from identifiable mechanisms rather than from generic accounting circularity.

The broader implication of our analysis is methodological. Empirical price-value comparisons should report not only distance measures, but also the features of the constructed system that help explain them: the dispersion of the direct labor vector, the column structure of the Leontief inverse, and the spectral properties of the input-output matrix. Sabatino’s paper is valuable because it makes the mechanics of the empirical procedure visible. Once those mechanics are examined, her artificial datasets do not invalidate price-value comparisons as tautological; they instead reveal structural determinants of price-value proximity that deserve closer attention.

A Functional dependence does not imply proximity

This appendix gives a simple numerical illustration of the distinction between functional dependence and proximity. Let

$$x = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \quad (20)$$

Now define another vector as a function of x :

$$y = f(x) = \begin{pmatrix} x_1 \\ x_2 \\ 100x_3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 300 \end{pmatrix}. \quad (21)$$

The vector y is entirely constructed from x , but it is not close to x . Using the angular distance,

$$d(x, y) = \sqrt{2(1 - \cos \theta)}, \quad \cos \theta = \frac{x'y}{\|x\|\|y\|}, \quad (22)$$

we obtain

$$\cos \theta = \frac{1 \cdot 1 + 2 \cdot 2 + 3 \cdot 300}{\sqrt{1^2 + 2^2 + 3^2} \sqrt{1^2 + 2^2 + 300^2}} = \frac{905}{\sqrt{14} \sqrt{90005}} \approx 0.806, \quad (23)$$

and therefore

$$d(x, y) \approx \sqrt{2(1 - 0.806)} \approx 0.623. \quad (24)$$

The example is deliberately simple. It shows that a vector can be fully determined by another vector without being close to it. To establish circularity in a proximity test, it is therefore not enough to show that one object is a function of another. One must show that the construction forces the relevant distance measure to be small.

B Dispersion and the angle-based distance

This appendix derives the relationship between the angle-based distance used in the paper and the coefficient of variation of the vector entering the distance calculation. Let

$$z = \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{pmatrix}$$

be a strictly positive vector, i.e. $z_i > 0$, $i = 1, 2, \dots, n$. In the application considered in the main text,

$$z_i = \frac{1}{\lambda_i},$$

where λ_i is the vertically integrated labor coefficient of sector i . Note that $\lambda_i > 0$ if A is productive. The distance statistic compares z to the unit vector

$$\mathbf{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}.$$

The angle-based distance is

$$d(z, \mathbf{1}) = \sqrt{2(1 - \cos \theta)}, \quad (25)$$

where θ is the angle between z and $\mathbf{1}$. By the standard dot-product formula,

$$\cos \theta(z, \mathbf{1}) = \frac{z' \mathbf{1}}{\|z\| \|\mathbf{1}\|}. \quad (26)$$

We now compute each term in this expression. The numerator is

$$z' \mathbf{1} = z_1 + z_2 + \cdots + z_n = \sum_{i=1}^n z_i. \quad (27)$$

Let $\bar{z}$ denote the mean of z :

$$\bar{z} = \frac{1}{n} \sum_{i=1}^n z_i. \quad (28)$$

It follows that

$$\sum_{i=1}^n z_i = n\bar{z}, \quad (29)$$

and therefore

$$z' \mathbf{1} = n\bar{z}. \quad (30)$$

The first term in the denominator is the Euclidean norm of z :

$$\|z\| = \sqrt{z_1^2 + z_2^2 + \cdots + z_n^2} = \sqrt{\sum_{i=1}^n z_i^2}. \quad (31)$$

The second term is the Euclidean norm of the unit vector:

$$\|\mathbf{1}\| = \sqrt{1^2 + 1^2 + \cdots + 1^2} = \sqrt{n}. \quad (32)$$

Thus,

$$\cos \theta(z, \mathbf{1}) = \frac{n\bar{z}}{\sqrt{\sum_{i=1}^n z_i^2} \sqrt{n}}. \quad (33)$$

We now rewrite $\sum_i z_i^2$ in terms of the mean and variance of z . Using the (uncorrected) sample variance,

$$\hat{\sigma}_z^2 = \frac{1}{n} \sum_{i=1}^n (z_i - \bar{z})^2. \quad (34)$$

Expanding the squared term gives

$$(z_i - \bar{z})^2 = z_i^2 - 2z_i\bar{z} + \bar{z}^2. \quad (35)$$

Summing over all sectors,

$$\sum_{i=1}^n (z_i - \bar{z})^2 = \sum_{i=1}^n z_i^2 - 2\bar{z} \sum_{i=1}^n z_i + \sum_{i=1}^n \bar{z}^2. \quad (36)$$

Since $\sum_i z_i = n\bar{z}$, the middle term becomes

$$-2\bar{z} \sum_{i=1}^n z_i = -2\bar{z}(n\bar{z}) = -2n\bar{z}^2. \quad (37)$$

Since $\bar{z}^2$ is constant across sectors,

$$\sum_{i=1}^n \bar{z}^2 = n\bar{z}^2. \quad (38)$$

Therefore,

$$\sum_{i=1}^n (z_i - \bar{z})^2 = \sum_{i=1}^n z_i^2 - n\bar{z}^2. \quad (39)$$

Dividing by n , we get

$$\hat{\sigma}_z^2 = \frac{1}{n} \sum_{i=1}^n z_i^2 - \bar{z}^2. \quad (40)$$

Rearranging,

$$\frac{1}{n} \sum_{i=1}^n z_i^2 = \hat{\sigma}_z^2 + \bar{z}^2. \quad (41)$$

Multiplying both sides by n ,

$$\sum_{i=1}^n z_i^2 = n(\hat{\sigma}_z^2 + \bar{z}^2). \quad (42)$$

The sample coefficient of variation of z is

$$\hat{C}V(z) = \frac{\hat{\sigma}_z}{\bar{z}}. \quad (43)$$

Equivalently,

$$\hat{\sigma}_z = \hat{C}V(z)\bar{z}, \quad (44)$$

and hence

$$\hat{\sigma}_z^2 = \hat{C}V(z)^2 \bar{z}^2. \quad (45)$$

Substituting this into the expression for $\sum_i z_i^2$, we obtain

$$\sum_{i=1}^n z_i^2 = n \left(\hat{C}V(z)^2 \bar{z}^2 + \bar{z}^2 \right) \quad (46)$$

$$= n\bar{z}^2 \left(1 + \hat{C}V(z)^2 \right). \quad (47)$$

Substituting this result into the expression for the cosine gives

$$\cos \theta(z, \mathbf{1}) = \frac{n\bar{z}}{\sqrt{n\bar{z}^2(1 + \hat{C}V(z)^2)}\sqrt{n}}. \quad (48)$$

Since z is positive, $\bar{z} > 0$, so

$$\sqrt{n\bar{z}^2(1 + \hat{C}V(z)^2)} = \sqrt{n}\bar{z}\sqrt{1 + \hat{C}V(z)^2}. \quad (49)$$

Therefore,

$$\cos \theta(z, \mathbf{1}) = \frac{n\bar{z}}{\sqrt{n}\bar{z}\sqrt{1 + \hat{C}V(z)^2}\sqrt{n}} \quad (50)$$

$$= \frac{n\bar{z}}{n\bar{z}\sqrt{1 + \hat{C}V(z)^2}} \quad (51)$$

$$= \frac{1}{\sqrt{1 + \hat{C}V(z)^2}}. \quad (52)$$

Finally, substituting the expression for $\cos \theta$ into the distance formula gives

$$d(z, \mathbf{1}) = \sqrt{2 \left(1 - \frac{1}{\sqrt{1 + \hat{C}V(z)^2}} \right)}. \quad (53)$$

This expression shows that the angle-based distance is fully determined by the coefficient of variation of z .

If

$$y = \left[1 - (1 + x^2)^{-1/2} \right]^{1/2}$$

then

$$\frac{dy}{dx} = \frac{\left[1 - (1 + x^2)^{-1/2} \right]^{-1/2}}{4(1 + x^2)^{3/2}} > 0.$$

In the limiting case where $CV(z) = 0$, all elements of z are identical, so z is proportional to $\mathbf{1}$, $\cos \theta = 1$, and

$$d(z, \mathbf{1}) = 0.$$

In the application in the main text, $z_i = 1/\lambda_i$. Hence,

$$d = \sqrt{2 \left(1 - \frac{1}{\sqrt{1 + CV(1/\lambda)^2}} \right)}. \quad (54)$$

The market-price/direct-price distance is therefore a monotone increasing function of the dispersion of $1/\lambda_i$, relative to its mean.

C Spectral radius and the smoothing of vertically integrated labor

If the $n \times n$ matrix A is primitive, that is, if it is nonnegative and irreducible with a single spectral radius, or maximal eigenvalue, r , where this eigenvalue is not a repeated root of the characteristic equation of A , then a theorem in Meyer (2000, p. 674) shows that

$$\lim_{k \rightarrow \infty} \frac{A^k}{r^k} = \frac{pq^T}{q^T p}, \quad (55)$$

where p and q are the $n \times 1$ Perron vectors, that is, the strictly positive vectors associated with the maximal eigenvalues and normalized to have the sum of their elements equal to unity, of A and A^T , respectively. Note that the matrix on the right-hand side of equation (55) is a rank-one matrix, that is, each column of the matrix is a multiple of the vector p . The implication is that for large k , the matrix A^k , that is, the matrix A multiplied by itself k times, starts behaving like a rank-one matrix with each column a scalar multiple of its Perron vector.

Now let us consider $(I - A)^{-1}$. When $r < 1$, then

$$(I - A)^{-1} = \sum_{k=0}^{\infty} A^k, \quad (56)$$

and the sum on the right-hand side is strictly positive, that is, each element of the matrix on the left is strictly positive. As $r \rightarrow 1$ from below, the sum grows monotonically in size and is increasingly dominated by terms involving higher powers of the matrix A , that is, A^k , where k is large. But equation (55) shows us that each of these matrices behaves like a rank-one matrix with each column a scalar multiple of its Perron vector. Hence, $(I - A)^{-1}$ approaches a matrix where all the columns are just scalar multiples of the Perron vector p .

Now consider the $1 \times n$ value vector

$$\lambda = l(I - A)^{-1}, \quad (57)$$

where l is the $1 \times n$ labor input vector. Note that, as $r \rightarrow 1$ from below, the relative variation in the elements of λ , captured for instance by the coefficient of variation, falls because each element of λ is just a scalar multiple of lp . This means that the variation of the vector composed of the reciprocals of the value vector also falls, and for the same reason.

D Construction of artificial datasets

This appendix describes how the artificial datasets used in the analysis are constructed in the replication code. The purpose is to make explicit which features of each construction are preserved and which are varied.

D.1 Scrambled datasets

The scrambled simulations begin from Sabatino's Scrambled workbook. The code reads three objects from the workbook: the input-output matrix A , the simple labor vector l , and the gross output vector Q . For each replication, the entries of these objects are independently reshuffled.

First, the entries of the input-output matrix are randomly permuted and then reshaped into a new matrix of the same dimension:

$$A^{(s)} = \text{shuffle}(A).$$

This preserves the empirical distribution of the original matrix entries, but breaks their original row and column locations. Second, the gross output vector is randomly permuted:

$$Q^{(s)} = \text{shuffle}(Q).$$

Third, the simple labor vector is randomly permuted:

$$l^{(s)} = \text{shuffle}(l).$$

For each scrambled draw, the code checks whether the resulting input-output system is viable. It records whether all column sums of $A^{(s)}$ are below one, whether the spectral radius satisfies $\rho(A^{(s)}) < 1$, and whether $(I - A^{(s)})^{-1}$ can be computed. Draws that fail these checks

are discarded. For valid draws, vertically integrated labor coefficients are computed as

$$\lambda^{(s)} = l^{(s)}(I - A^{(s)})^{-1}.$$

Sabatino direct prices are then constructed from $\lambda^{(s)}$ and $Q^{(s)}$, and the code computes the distance measures, including the angle-based D -distance and $CV(1/\lambda)$. The scrambled exercise therefore preserves the marginal distributions of the original inputs while breaking their sectoral correspondence.

D.2 Random Numbers datasets

The Random Numbers simulations do not use an Excel workbook as an input. Instead, the code reconstructs Sabatino’s random data-generating process directly. Each draw contains 65 sectors.

The use table U is generated by drawing all 65×65 entries from a lognormal distribution:

$$U_{ij} \sim \text{Lognormal}(8, 2).$$

Let

$$u_i = \sum_j U_{ji}$$

denote the intermediate input total of industry i , that is, the column sum of the use table.

D.2.1 Fixed-share version

In the fixed-share version, the components of value added are fixed proportions of industry intermediate input:

$$c_i = 0.01u_i, \quad t_i = 0.10u_i, \quad \pi_i = 0.15u_i,$$

where c_i is compensation, t_i is taxes, and π_i is profits. Value added is

$$va_i = c_i + t_i + \pi_i,$$

and industry gross output is

$$g_i = u_i + va_i.$$

Because compensation is fixed at one percent of intermediate input, the simple labor vector is compressed by construction:

$$l_i = \frac{c_i}{g_i} = \frac{0.01u_i}{u_i + 0.26u_i} = \frac{0.01}{1.26}.$$

Thus, in the fixed-share version, the simple labor vector is exactly constant across sectors.

Commodity gross output is constructed as the row sum of the use table plus final demand. In the code, the vector of final-demand row sums is set equal to value added, so

$$q_i = \sum_j U_{ij} + va_i.$$

The make table V is then generated to match the industry and commodity totals. To construct a feasible random make table, our simulation sets the diagonal entry in each sector equal to a random fraction of the smaller of the corresponding industry and commodity totals:

$$V_{ii} = \sigma_i \min(g_i, q_i), \quad \sigma_i \sim U(0.01, 0.99).$$

The use of the minimum ensures that the diagonal entry does not exceed either the relevant industry total or the relevant commodity total, so that the remaining row and column totals are nonnegative. The remaining totals are then distributed across off-diagonal entries so that the make table satisfies the required industry and commodity totals.

Finally, the input-output matrix is constructed as

$$A = V\hat{q}^{-1}U\hat{g}^{-1},$$

where $\hat{q}$ and $\hat{g}$ are diagonal matrices with q and g on the diagonal. The draw is retained only if $\rho(A) < 1$. For valid draws, the code computes

$$\lambda = l(I - A)^{-1},$$

constructs Sabatino direct prices, and records the distance measures.

D.2.2 Heterogeneous-share version

The heterogeneous-share version preserves the aggregate shares in Sabatino's random construction but allows sectoral income shares to vary. Define sector-specific shares relative to intermediate input:

$$s_i^c = \frac{c_i}{u_i}, \quad s_i^t = \frac{t_i}{u_i}, \quad s_i^\pi = \frac{\pi_i}{u_i}.$$

Instead of fixing these shares at 0.01, 0.10, and 0.15 in every sector, the code draws preliminary shares from lognormal distributions:

$$\tilde{s}_i^c \sim \text{Lognormal}(0, 1.5),$$

$$\tilde{s}_i^t \sim \text{Lognormal}(0, 1.0),$$

$$\tilde{s}_i^\pi \sim \text{Lognormal}(0, 1.0).$$

These preliminary shares are then rescaled so that the aggregate shares relative to total intermediate input match Sabatino’s values. Let

$$\omega_i = \frac{u_i}{\sum_h u_h}.$$

For each component $k \in \{c, t, \pi\}$, define

$$s_i^k = \bar{s}^k \frac{\tilde{s}_i^k}{\sum_j \omega_j \tilde{s}_j^k},$$

where

$$\bar{s}^c = 0.01, \quad \bar{s}^t = 0.10, \quad \bar{s}^\pi = 0.15.$$

This normalization ensures

$$\frac{\sum_i s_i^k u_i}{\sum_i u_i} = \bar{s}^k.$$

The heterogeneous-share design therefore preserves the same aggregate accounting proportions as the fixed-share design, but removes the sector-by-sector equality of income shares. The remaining steps are the same as in the fixed-share version: construct value added, gross output, commodity gross output, the make table, A , λ , Sabatino direct prices, and the distance measures.

D.3 HUMBUG datasets

The HUMBUG simulations begin from Sabatino’s HUMBUG workbook. The code reads the pre-HUMBUG core use table from the sheet `UseTableGovernmentMerger` and the workbook HUMBUG use table from `UseTableReconstructed`. The core use table is a 65×65 block. The HUMBUG mask is detected by comparing the reconstructed HUMBUG use table with Sabatino’s inserted value:

$$h = 11,111,111,111.$$

Cells equal to this value inside the HUMBUG footprint are treated as part of the HUMBUG mask.

The code also reads value added, compensation, gross output, commodity gross output, employment, and the reconstructed make table from the workbook. For any constructed use table variant, industry gross output is recomputed as

$$g_i = \sum_j U_{ji} + va_i,$$

and commodity gross output is recomputed as

$$q_i = \sum_j U_{ij} + f_i,$$

where f_i is the final-use component implied by the workbook. The make table is then reconstructed so that its row and column totals match the new g and q , using the workbook's diagonal make-table shares. The resulting input-output matrix is computed from the use and make tables as

$$A = B_U D_V,$$

where

$$B_U = U \hat{g}^{-1}, \quad D_V = V \hat{q}^{-1}.$$

The simple labor vector is recomputed for each variant using compensation, employment, and the new industry gross output. The code then computes $\lambda = l(I - A)^{-1}$, Sabatino direct prices, D -distance, $CV(\lambda)$, $CV(1/\lambda)$, $\rho(A)$, and column-sum diagnostics.

D.3.1 Column-heterogeneity HUMBUG variants

The first HUMBUG family tests whether the column distribution of the HUMBUG entries drives the low distance. The starting point is the pre-HUMBUG use table and the same HUMBUG mask as in the workbook. In Sabatino's construction, every cell in the mask receives the same inserted value h . The total amount artificially inserted into the table is therefore

$$h \times N_H,$$

where N_H is the number of cells in the HUMBUG mask.

In the column-heterogeneity variants, this total inserted amount is kept fixed, but it is distributed unevenly across columns. For each column j of the HUMBUG mask, the code

draws a multiplier

$$z_j \sim \text{Lognormal}(0, \sigma),$$

where

$$\sigma \in \{0.25, 0.50, 1.00, 1.50, 2.00\}.$$

Before rescaling, the value inserted into each HUMBUG cell in column j would be hz_j . The code then rescales all column multipliers by a common factor so that the sum of all inserted HUMBUG entries remains equal to the workbook total hN_H . If n_j is the number of HUMBUG cells in column j , the unscaled inserted total is

$$h \sum_j n_j z_j.$$

The rescaling factor is therefore

$$\kappa = \frac{hN_H}{h \sum_j n_j z_j} = \frac{N_H}{\sum_j n_j z_j}.$$

The value inserted into each HUMBUG cell in column j is then

$$h_j = h\kappa z_j.$$

Thus, some HUMBUG columns receive larger artificial entries and others receive smaller artificial entries, but the total amount inserted across all HUMBUG cells is unchanged. This isolates the effect of column distribution from the overall scale of the intervention.

D.3.2 Spectral-radius HUMBUG variants

The second HUMBUG family tests whether the scale of the inserted values drives the low distance. The same HUMBUG mask is used, and each cell in the mask receives the same inserted value, as in Sabatino's construction. However, instead of fixing the value at $h = 11,111,111,111$, the code chooses values that produce target spectral radii:

$$\rho(A) \in \{0.50, 0.60, 0.70, 0.80, 0.90, 0.95, 0.98, 0.99, 0.995, 0.997, 0.9975\}.$$

For each target, the code searches for the common HUMBUG-cell value that makes the resulting input-output matrix attain the desired spectral radius. This creates a sequence of use tables that all retain the same HUMBUG mask, but differ in the intensity of the inserted values and therefore in how close the production system is to the productivity boundary.

These two HUMBUG families separate the column-pattern channel from the spectral-

radius channel. The column-heterogeneity variants change the distribution of inserted values across HUMBUG columns while holding the total inserted amount fixed. The spectral-radius variants preserve the HUMBUG pattern while changing the overall size of the inserted values.

References

- Basu, D. (2017). “Quantitative Empirical Research in Marxist Political Economy: A Selective Review”. In: *Journal of Economic Surveys* 31.5, pp. 1359–1386. ISSN: 1467-6419. DOI: 10.1111/joes.12218. (Visited on 07/26/2022).
- Ferrer-Hernández, J. and L. D. Torres-González (2021). “Eigenvalues and Eigenlabors: On Iliadi’s, Mariolis’, Soklis’, and Tsoulfidis’ Explanation of the Empirical Regularities in Price Curves”. In: *Working Papers* 2119. (Visited on 07/18/2025).
- İşıkara, G. and P. Mokre (2022). “Price-Value Deviations and the Labour Theory of Value: Evidence from 42 Countries, 2000–2017”. In: *Review of Political Economy* 34.1, pp. 165–180. ISSN: 0953-8259. DOI: 10.1080/09538259.2021.1904648. (Visited on 07/31/2024).
- Kliman, A. J. (2002). “The Law of Value and Laws of Statistics: Sectoral Values and Prices in the US Economy, 1977—97”. In: *Cambridge Journal of Economics* 26.3, pp. 299–311. ISSN: 0309-166X. JSTOR: 23600287. (Visited on 07/31/2024).
- Meyer, M. (2000). “What Is Special about Patent Citations? Differences between Scientific and Patent Citations”. In: *Scientometrics* 49.1, pp. 93–123. ISSN: 1588-2861. DOI: 10.1023/A:1005613325648. (Visited on 03/21/2025).
- Moraitis, T. and D. Basu (2026). “Alternative Approaches to Labor Values and Prices of Production: Theory and Evidence”. In: *Structural Change and Economic Dynamics* 78, pp. 171–187. ISSN: 0954-349X. DOI: 10.1016/j.strueco.2026.03.009. (Visited on 04/24/2026).
- Ochoa, E. M. (1984). “Labor-Value and Prices of Production: An Interindustry Study of the U.S. Economy, 1947–1972”. PhD thesis. New York: New School for Social Research. (Visited on 07/31/2024).
- Sabatino, I. (2026). “Humbug Labor Values”. In: *Capital As Power*. (Visited on 04/24/2026).
- Shaikh, A. (1974). “Laws of Production and Laws of Algebra: The Humbug Production Function”. In: *The Review of Economics and Statistics* 56.1, pp. 115–120. ISSN: 0034-6535. DOI: 10.2307/1927538. JSTOR: 1927538. (Visited on 06/03/2026).
- (1984). “The Transformation from Marx to Sraffa”. In: *Ricardo, Marx, Sraffa: The Langston Memorial Volume*. London, UK: Verso, pp. 43–84. (Visited on 07/31/2024).
- Steedman, I. and J. Tomkins (1998). “On Measuring the Deviation of Prices from Values”. In: *Cambridge Journal of Economics* 22.3, pp. 379–385. ISSN: 0309-166X. JSTOR: 23600308. (Visited on 07/31/2024).
- Tsoulfidis, L. (2002). “Values, Prices of Production and Market Prices: Some More Evidence from the Greek Economy”. In: *Cambridge Journal of Economics* 26.3, pp. 359–369. ISSN: 0309-166X. (Visited on 07/31/2024).

Tsoufidis, L. and D. M. Rieu (2006). “Labor Values, Prices of Production, and Wage-Profit Rate Frontiers of the Korean Economy”. In: 19.3, pp. 275–296. issn: 1225-0279. HDL: 10371/1355.